附录 B 概率论与数理统计

概率论与数理统计是金融计量的重要数学基础，本附录汇总了书中涉及到的相关概念和结果，主要内容包括：随机序列的收敛方式、时间序列的渐近理论、Bootstrap 的渐近理论、多元正态分布及其性质和常用概率不等式。进一步深入学习概率论与数理统计的基础知识可以参考 Hansen (2022b) 和苏淳，冯群强 (2020)；李高荣，吴密霞 (2021) 对多元统计分析进行了较全面地介绍；Lin et al. (2010) 总结了大多数常用的概率不等式；时间序列的渐近理论可以参考 Hall et al. (1980)。

B.1 矩生成函数

顾名思义，矩生成函数（moment generating function, mgf）可以用来计算随机变量的各阶矩。然而，计算矩并不是 mgf 的唯一用途，它的另一大用途是刻画概率分布，通过随机变量的 mgf 的具体形式，可以判断随机变量服从何种分布。下面给出 mgf 的定义：

定义 B.1. 矩生成函数

令 x 是随机变量，其分布函数为 F_{x}，x 的矩生成函数，或称为 F_{x} 的矩生成函数，定义为：

$$M_{\mathrm{x}}(t) = \mathbb{E}\left[\exp(t\mathrm{x})\right].$$

如果存在 $\delta > 0$ 使得，对所有的 $-\delta < t < \delta$，$\mathbb{E}\left[\exp(t\mathrm{x})\right]$ 存在，则称 x 的矩生成函数存在；否则，称 x 的矩生成函数不存在。♣

矩生成函数具有如下重要的性质：

性质（矩生成性质）如果 x 的 mgf 为 $M_{\mathrm{x}}(t)$，则：

$$\mathbb{E}(\mathrm{x}^k) = M_{\mathrm{x}}^{(k)}(0),$$

其中，

$$M_{\mathrm{x}}^{(k)}(0) \equiv \left.\frac{d^k}{dt^k} M_{\mathrm{x}}(t)\right|_{t=0},$$

即，x 的 k 阶矩等于 $M_{\mathrm{x}}(t)$ 的 k 阶导在 $t = 0$ 处的取值。（证明见 Casella et al. (2002) 的定理 2.3.7）

性质（唯一性）令 $F_{\mathrm{x}}(x)$ 和 $F_{\mathrm{y}}(y)$ 是两个累积分布函数，并且它们的各阶矩存在，则有：

(a) 如果 x 和 y 具有有限的支撑，则 $F_{\mathrm{x}}(u) = F_{\mathrm{y}}(u)$，$\forall\, u$，当且仅当 $\mathbb{E}(\mathrm{x}^r) = \mathbb{E}(\mathrm{y}^r)$，$r = 0, 1, 2, \ldots$；

(b) 若 x 和 y 的 mgf 存在，并且，对 0 的某一邻域内的所有 t，$M_{\mathrm{x}}(t) = M_{\mathrm{y}}(t)$，则 $F_{\mathrm{x}}(u) = F_{\mathrm{y}}(u)$，$\forall\, u$。

（证明见 Shao (2003) 的定理 1.6）

性质（收敛性质）令 $\{\mathrm{x}_i,\ i = 1, 2, \ldots\}$ 是随机变量的序列，x_i 的 mgf 为 $M_{\mathrm{x}_i}(t)$。若存在

$\delta > 0$ 使得:

$$\lim_{i\to\infty} M_{\mathrm{x}_i}(t) = M_{\mathrm{x}}(t), \quad \forall\, t \in (-\delta, \delta),$$

其中, $M_{\mathrm{x}}(t)$ 是一 mgf,则存在唯一的累积分布函数 F_{x},其各阶矩由 $M_{\mathrm{x}}(t)$ 确定,并且,对于所有满足 $F_{\mathrm{x}}(x)$ 连续的 x,我们有:

$$\lim_{i\to\infty} F_{\mathrm{x}_i}(x) = F_{\mathrm{x}}(x).$$

(证明见 Feller (1971) 的 433 页定理 2a)

性质 (仿射变换的矩生成函数)对任意的常数 a 和 b,随机变量 $a\mathrm{x}+b$ 的 mgf 为:

$$M_{a\mathrm{x}+b}(t) = \exp(bt) M_{\mathrm{x}}(at).$$

(证明见 Casella et al. (2002) 的定理 2.3.15)

B.2 变换

令 x 是随机变量,其 cdf 为 $F_{\mathrm{x}}(x)$,样本空间为 $\mathcal{X}$,x 的任意函数 $\mathrm{y} = g(\mathrm{x})$ 也是随机变量,$g(\cdot)$ 定义了从 $\mathcal{X}$ 到新的样本空间 $\mathcal{Y}$ 的一个映射。随机变量 y 的分布由 $F_{\mathrm{x}}(x)$ 和 $g(\cdot)$ 确定:

$$\mathbb{P}\left[\mathrm{y} \in A\right] = \mathbb{P}\left[g(\mathrm{x}) \in A\right] = \mathbb{P}\left[\mathrm{x} \in g^{-1}(A)\right],$$

其中,$g^{-1}(A) = \{x \in \mathcal{X} \mid g(x) \in A\}$。

如果 x 是离散随机变量,则 $\mathcal{X}$ 是可数集,$\mathcal{Y} = \{y \mid y = g(x),\ x \in \mathcal{X}\}$ 也是可数集。因此,y 也是离散随机变量,其概率质量(mass)函数为:

$$f_{\mathrm{y}}(y) = \mathbb{P}(\mathrm{y} = y) = \sum_{x \in g^{-1}(y)} \mathbb{P}(\mathrm{x} = x), \quad \forall\, y \in \mathcal{Y}.$$

如果 x 和 y 是连续随机变量,则 y 的 cdf 为:

$$F_{\mathrm{y}}(y) = \mathbb{P}[g(\mathrm{x}) \leqslant y] = \int_{\{x \in \mathcal{X} \mid g(x) \leqslant y\}} f_{\mathrm{x}}(x) dx.$$

当 $g(\cdot)$ 为单调函数时,有如下结论:(证明过程见 Casella et al. (2002) 的第 50 页)

定理 B.1. 单调变换的累积分布函数

设 $\mathrm{x} \sim F_{\mathrm{x}}(x)$,$\mathrm{y} = g(\mathrm{x})$,定义:

$$\mathcal{X} = \{x \mid f_{\mathrm{x}}(x) > 0\}, \quad \mathcal{Y} = \{y \mid \text{存在 } x \in \mathcal{X} \text{ 使得 } y = g(x)\}. \tag{B.1}$$

(a) 如果 g 是 $\mathcal{X}$ 上的严格单调递增函数,则,对任意的 $y \in \mathcal{Y}$, $F_{\mathrm{y}}(y) = F_{\mathrm{x}}(g^{-1}(y))$;

(b) 如果 g 是 $\mathcal{X}$ 上的严格单调递减函数且 x 是连续随机变量,则,对任意的 $y \in \mathcal{Y}$, $F_{\mathrm{y}}(y) = 1 - F_{\mathrm{x}}(g^{-1}(y))$。

考虑区间 $(x, x+dx)$,其中,dx 是一个微小的增量,函数 $g(\cdot)$ 将此区间映射为 $(y, y+dy)$,容易知道这两个区间的概率质量应该相同,即,$f_{\mathrm{x}}(x)dx = f_{\mathrm{y}}(y)dy$,等价地,

$$f_{\mathrm{y}}(y) = f_{\mathrm{x}}(x)\frac{dx}{dy}.$$

事实上,由定理 B.1 和求导链式法则,可以推出如下重要结论:

定理 B.2. 单调变换的概率密度函数

设 $\mathrm{x} \sim F_{\mathrm{x}}(x)$，$\mathrm{y} = g(\mathrm{x})$，其中，$g(\cdot)$ 是严格单调函数。并且，$f_{\mathrm{x}}(x)$ 在 $\mathcal{X}$ 上连续，$g^{-1}(y)$ 在 $\mathcal{Y}$ 上具有连续导数，其中，$\mathcal{X}$ 和 $\mathcal{Y}$ 的定义与 (B.1) 相同，则随机变量 y 的概率密度函数为：

$$f_{\mathrm{y}}(y) = \begin{cases} f_{\mathrm{x}}\left(g^{-1}(y)\right)\left|\dfrac{d}{dy}g^{-1}(y)\right|, & y \in \mathcal{Y} \\ 0, & y \notin \mathcal{Y}. \end{cases} \tag{B.2}$$

♡

上述结论可以推广到多元情形：设 $\mathbf{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_n)'$ 是 n 维随机向量，其联合密度为 $f_{\mathbf{x}}(\boldsymbol{x})$。进一步，再假设 $\mathbf{y} = \boldsymbol{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_n(\mathbf{x}))'$ 是一一对应（one-to-one correspondence）的变换，并且逆变换 $\mathbf{x} = \boldsymbol{g}^{-1}(\mathbf{y})$ 存在，若偏导数 $\frac{\partial x_i}{\partial y_j}$，$(i, j = 1, 2, \ldots, n)$，存在且连续，则 $\mathbf{y} = \boldsymbol{g}(\mathbf{x})$ 的联合密度函数存在，并且可以表达为：

$$f_{\mathbf{y}}(\boldsymbol{y}) = f_{\mathbf{x}}(\boldsymbol{g}^{-1}(\boldsymbol{y})) \cdot |\det \boldsymbol{J}|,$$

其中，

$$[\boldsymbol{J}]_{ij} = \frac{\partial x_i}{\partial y_j}, \quad i, j = 1, 2, \ldots, n,$$

称 $\det \boldsymbol{J}$ 为变换 $\mathbf{y} = \boldsymbol{g}(\mathbf{x})$ 的雅可比行列式。

B.3 随机向量和随机矩阵

若一向量的元素为随机变量，则称此向量为随机向量。类似地，若一矩阵的元素为随机变量，则称此矩阵为随机矩阵。令 $\mathbf{X} = (\mathrm{x}_{ij})$ 为 $m \times n$ 维随机矩阵，$\mathbf{X}$ 的期望（假设期望存在），记为 $\mathbb{E}[\mathbf{X}]$，定义为：

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}(\mathrm{x}_{11}) & \mathbb{E}(\mathrm{x}_{12}) & \cdots & \mathbb{E}(\mathrm{x}_{1n}) \\ \mathbb{E}(\mathrm{x}_{21}) & \mathbb{E}(\mathrm{x}_{22}) & \cdots & \mathbb{E}(\mathrm{x}_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}(\mathrm{x}_{m1}) & \mathbb{E}(\mathrm{x}_{n2}) & \cdots & \mathbb{E}(\mathrm{x}_{mn}) \end{bmatrix}.$$

特别地，当 $n = 1$ 时，随机向量 $\mathbf{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_m)'$ 的期望为：

$$\mathbb{E}[\mathbf{x}] = \begin{bmatrix} \mathbb{E}(\mathrm{x}_1) \\ \mathbb{E}(\mathrm{x}_2) \\ \vdots \\ \mathbb{E}(\mathrm{x}_m) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix} \equiv \boldsymbol{\mu}.$$

随机矩阵的期望具有如下线性性质：令 $\mathbf{X}$ 和 $\mathbf{Y}$ 为随机矩阵，常数矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 满足可相乘条件，则有：

$$\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}[\mathbf{X}] + \mathbb{E}[\mathbf{Y}]; \tag{B.3}$$

$$\mathbb{E}(\boldsymbol{A}\mathbf{X}\boldsymbol{B}) = \boldsymbol{A}\mathbb{E}[\mathbf{X}]\boldsymbol{B}. \tag{B.4}$$

随机向量 $\mathbf{x}$ 的协方差矩阵定义为：$\mathrm{Var}(\mathbf{x}) \equiv \mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'\right]$。具体地，

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'\right] \\
&= \mathbb{E}\left(\begin{bmatrix} \mathrm{x}_1-\mu_1 \\ \mathrm{x}_2-\mu_2 \\ \vdots \\ \mathrm{x}_m-\mu_m \end{bmatrix}[\mathrm{x}_1-\mu_1, \mathrm{x}_2-\mu_2, \ldots, \mathrm{x}_m-\mu_m]\right) \\
&= \begin{bmatrix}
\mathbb{E}\left(\mathrm{x}_1-\mu_1\right)^2 & \mathbb{E}\left(\mathrm{x}_1-\mu_1\right)\left(\mathrm{x}_2-\mu_2\right) & \cdots & \mathbb{E}\left(\mathrm{x}_1-\mu_1\right)\left(\mathrm{x}_m-\mu_m\right) \\
\mathbb{E}\left(\mathrm{x}_2-\mu_2\right)\left(\mathrm{x}_1-\mu_1\right) & \mathbb{E}\left(\mathrm{x}_2-\mu_2\right)^2 & \cdots & \mathbb{E}\left(\mathrm{x}_2-\mu_2\right)\left(\mathrm{x}_m-\mu_m\right) \\
\vdots & \vdots & \ddots & \vdots \\
\mathbb{E}\left(\mathrm{x}_m-\mu_m\right)\left(\mathrm{x}_1-\mu_1\right) & \mathbb{E}\left(\mathrm{x}_m-\mu_m\right)\left(\mathrm{x}_2-\mu_2\right) & \cdots & \mathbb{E}\left(\mathrm{x}_m-\mu_m\right)^2
\end{bmatrix} \\
&= \begin{bmatrix}
\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\
\sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{m1} & \sigma_{m2} & \cdots & \sigma_m^2
\end{bmatrix}.
\end{aligned}
$$

因为 $\sigma_{ij} = \mathbb{E}\left(\mathrm{x}_i-\mu_i\right)\left(\mathrm{x}_j-\mu_j\right) = \mathbb{E}\left(\mathrm{x}_j-\mu_j\right)\left(\mathrm{x}_i-\mu_i\right) = \sigma_{ji}$，所以 $\boldsymbol{\Sigma}$ 是对称矩阵。同时，容易验证：$\boldsymbol{\Sigma}$ 是半正定矩阵。

随机向量的期望和协方差矩阵具有如下性质：

- 令 $\boldsymbol{c}$ 为 m 维常数向量，则有：

$$\mathbb{E}(\boldsymbol{c}'\mathbf{x}) = \boldsymbol{c}'\boldsymbol{\mu}, \quad \mathrm{Var}(\boldsymbol{c}'\mathbf{x}) = \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}. \tag{B.5}$$

- 令 $\mathbf{z} = \boldsymbol{C}\mathbf{x}$，其中，$\boldsymbol{C}$ 为 $n \times m$ 维常数矩阵，则有：

$$\mathbb{E}(\mathbf{z}) = \boldsymbol{C}\boldsymbol{\mu}, \quad \mathrm{Var}(\mathbf{z}) = \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'. \tag{B.6}$$

- 令 $\boldsymbol{A}$ 是 $m \times m$ 维常数矩阵，则有：

$$\mathbb{E}(\mathbf{x}'\boldsymbol{A}\mathbf{x}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}. \tag{B.7}$$

令 $\mathbf{x}$ 和 $\mathbf{y}$ 分别是 m 维和 n 维随机向量，若对于任意的 $1 \leqslant i \leqslant m$，$1 \leqslant j \leqslant n$，$\mathrm{x}_i$ 和 y_j 的协方差 $\mathrm{Cov}(\mathrm{x}_i, \mathrm{y}_j)$ 存在，则称

$$\mathrm{Cov}(\mathbf{x}, \mathbf{y}) \equiv \mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{y}})'\right]$$

为随机向量 $\mathbf{x}$ 和 $\mathbf{y}$ 的协方差矩阵，其中，$\boldsymbol{\mu}_{\mathbf{x}}$ 和 $\boldsymbol{\mu}_{\mathbf{y}}$ 分别是 $\mathbf{x}$ 和 $\mathbf{y}$ 的期望向量。简单计算可知 $[\mathrm{Cov}(\mathbf{x}, \mathbf{y})]_{ij} = \mathrm{Cov}(\mathrm{x}_i, \mathrm{y}_j)$。两个随机向量的协方差矩阵具有如下性质：令常数矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 满足可相乘条件，则有：

$$\mathrm{Cov}(\boldsymbol{A}\mathbf{x}, \boldsymbol{B}\mathbf{y}) = \boldsymbol{A}\,\mathrm{Cov}(\mathbf{x}, \mathbf{y})\boldsymbol{B}'.$$

B.4 渐近理论

渐近理论是分析估计量和检验统计量大样本性质的理论基础，它首先涉及的是随机变量序列的不同收敛方式。令随机变量的序列 $\{\mathrm{x}_n,\ n \in \mathbb{N}\}$ 和随机变量 x 定义在同一个

概率空间 $(\boldsymbol{\Omega}, \mathscr{F}, \mathbb{P})$ 上，如果，对任意的 $\varepsilon > 0$，

$$\lim_{n\to\infty} \mathbb{P}(|\mathrm{x}_n - \mathrm{x}| > \varepsilon) = 0,$$

则称 $\{\mathrm{x}_n\}$ 依概率收敛（convergence in probability）到随机变量 x，记为 $\operatorname{plim}_{n\to\infty} \mathrm{x}_n = \mathrm{x}$，或者，$\mathrm{x}_n \underset{p}{\to} \mathrm{x}$。类似地，设 $\{\mathbf{x}_n\}$ 是 K 维随机向量的序列，$\mathbf{x}$ 是 K 维随机向量，若对任意的 $k = 1, \ldots, K$，$\{\mathrm{x}_{nk}\}$ 依概率收敛到 x_k，其中，x_{nk} 和 x_k 分别表示 $\mathbf{x}_n$ 和 $\mathbf{x}$ 的第 k 个元素，则称 $\{\mathbf{x}_n\}$ 依概率收敛到 $\mathbf{x}$。

如果 $\{\mathrm{x}_n, n \in \mathbb{N}\}$ 和 x 满足：

$$\mathbb{P}\left\{\omega \mid \lim_{n\to\infty} \mathrm{x}_n(\omega) = \mathrm{x}(\omega)\right\} = 1,$$

则称 $\{\mathrm{x}_n, n \in \mathbb{N}\}$ 以概率 1 收敛到 x，或，几乎必然（almost sure）收敛到 x，记为 $\mathrm{x}_n \underset{a.s.}{\to} \mathrm{x}$。

如果 $\{\mathrm{x}_n, n \in \mathbb{N}\}$ 和 x 满足：对 $r > 0$，有：

$$\lim_{n\to\infty} \mathbb{E}\left|\mathrm{x}_n - \mathrm{x}\right|^r = 0,$$

则称 $\{\mathrm{x}_n, n \in \mathbb{N}\}$ 依 r 阶平均（r-th mean）收敛到 x，或，L_r 收敛到 x，记为 $\mathrm{x}_n \underset{L_r}{\to} \mathrm{x}$。当 $r = 2$ 时，称 $\{\mathrm{x}_n, n \in \mathbb{N}\}$ 依均方（mean square）收敛到 x，记为 $\mathrm{x}_n \underset{m.s.}{\to} \mathrm{x}$。以概率 1 收敛和依 r 阶平均收敛都可以按照依概率收敛的方式推广到随机向量的情形。

令 $\{\mathrm{x}_n, n \in \mathbb{N}\}$ 是随机标量序列满足 $\mathrm{x}_n \sim F_n(x)$，随机变量 x 的分布函数为 $F(x)$（x 和 $\{\mathrm{x}_n, n \in \mathbb{N}\}$ 可以具有不同的概率空间），如果

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \text{对任意的 } x \text{ 满足 } F(x) \text{ 连续},$$

则称 $\{\mathrm{x}_n, n \in \mathbb{N}\}$ 依分布收敛到 x，或，依法则（convergence in law）收敛到 x，并且，称 F 为 x_n 的渐近分布，记为 $\mathrm{x}_n \underset{d}{\to} \mathrm{x}$，或，$\mathrm{x}_n \underset{L}{\to} \mathrm{x}$。

若 $\{\mathbf{x}_n\}$ 是随机向量，逐个元素的依分布收敛不一定能推出随机向量的依分布收敛。然而，我们有如下 Cramér-Wold 定理（证明见 Billingsley (1995) 的定理 29.4）：

- 令 $\{\mathbf{x}_n\}$ 是 K 维随机向量的序列，则有：

$$\mathbf{x}_n \underset{d}{\to} \mathbf{x} \;\Leftrightarrow\; \boldsymbol{\lambda}'\mathbf{x}_n \underset{d}{\to} \boldsymbol{\lambda}'\mathbf{x}, \quad \forall\, \boldsymbol{\lambda} \in \mathbb{R}^K.$$

上述四种不同的收敛方式之间具有如下关系：

- L_r 收敛蕴涵依概率收敛，反之不成立；
- 如果 $\mathrm{x}_n \underset{p}{\to} \mathrm{x}$，则有 $\mathrm{x}_n \underset{d}{\to} \mathrm{x}$，反之不成立；
- $\mathrm{x}_n \underset{p}{\to} c$ 等价于 $\mathrm{x}_n \underset{d}{\to} c$，其中，$c$ 为常数；
- 如果 $\mathrm{x}_n \underset{a.s.}{\to} \mathrm{x}$，则有 $\mathrm{x}_n \underset{p}{\to} \mathrm{x}$，反之不成立；
- L_r 收敛与几乎必然收敛互不蕴涵。

苏淳，冯群强 (2020) 的第六章给出了上述关系的证明和不成立的逆命题的反例。

在第 2.6 节中给出的最小二乘估计量的一致性是指最小二乘估计量依概率收敛到待估参数的真实值。一般地，令 $\hat{\boldsymbol{\theta}}_n$ 为参数向量 $\boldsymbol{\theta}$ 的估计量，下标 n 表示 $\hat{\boldsymbol{\theta}}_n$ 是基于样本量为 n 的样本计算得到，如果 $\hat{\boldsymbol{\theta}}_n \underset{p}{\to} \boldsymbol{\theta}$，则称 $\hat{\boldsymbol{\theta}}_n$ 是 $\boldsymbol{\theta}$ 的一致估计量。

收敛序列的连续映射具有如下性质：（证明见 Van der Vaart (2000) 的定理 2.3）

定理 B.3. 连续映射定理（Continuous Mapping Theorem）

令 $\boldsymbol{g}:\mathbb{R}^K\to\mathbb{R}^M$ 在集合 C 上的每一点连续，集合 C 满足 $\mathbb{P}(\mathrm{x}\in C)=1$，

- 如果 $\mathbf{x}_n\underset{d}{\to}\mathbf{x}$，则 $\boldsymbol{g}(\mathbf{x}_n)\underset{d}{\to}\boldsymbol{g}(\mathbf{x})$；
- 如果 $\mathbf{x}_n\underset{p}{\to}\mathbf{x}$，则 $\boldsymbol{g}(\mathbf{x}_n)\underset{p}{\to}\boldsymbol{g}(\mathbf{x})$；
- 如果 $\mathbf{x}_n\underset{a.s.}{\to}\mathbf{x}$，则 $\boldsymbol{g}(\mathbf{x}_n)\underset{a.s.}{\to}\boldsymbol{g}(\mathbf{x})$。

如下性质将依概率收敛和依分布收敛相结合，在推导估计量的大样本性质时经常会被用到：

定理 B.4. Slutsky 定理

令 $\mathbf{x}_n$ 和 $\mathbf{y}_n$ 是 K 维随机向量，如果 $\mathbf{x}_n\underset{d}{\to}\mathbf{x}$，$\mathbf{y}_n\underset{p}{\to}\boldsymbol{\alpha}$，其中，$\boldsymbol{\alpha}$ 为常数向量，则有 $\mathbf{x}_n+\mathbf{y}_n\underset{d}{\to}\mathbf{x}+\boldsymbol{\alpha}$；此外，令 $\mathbf{A}_n$ 为 $M\times K$ 维随机矩阵，满足 $\mathbf{A}_n\underset{p}{\to}\mathbf{A}$（可以理解为 $\mathrm{vec}\,(\mathbf{A}_n)\underset{p}{\to}\mathrm{vec}\,(\mathbf{A})$），则有 $\mathbf{A}_n\mathbf{x}_n\underset{d}{\to}\mathbf{A}\mathbf{x}$。

除了 Slutsky 定理外，容易验证以下两个结论也是成立的：

- 如果 $\mathbf{x}_n\underset{d}{\to}\mathbf{x}$，$\mathbf{y}_n\underset{p}{\to}\mathbf{0}$，则有 $\mathbf{y}_n'\mathbf{x}_n\underset{p}{\to}0$；
- 如果 $\mathbf{x}_n\underset{d}{\to}\mathbf{x}$，$\mathbf{A}_n\underset{p}{\to}\mathbf{A}$，其中，$\mathbf{A}$ 非奇异，则有：

$$\mathbf{x}_n'\mathbf{A}_n^{-1}\mathbf{x}_n\underset{d}{\to}\mathbf{x}'\mathbf{A}^{-1}\mathbf{x}. \tag{B.8}$$

若已知 $\mathbf{x}_n$ 的渐近分布，其非线性一阶连续可微函数的渐近分布可以通过如下方法得到：（证明见 Hayashi (2000) 的引理 2.5）

定理 B.5. Delta 方法

令 $\{\mathbf{x}_n\}$ 是 K 维随机向量的序列满足 $\mathbf{x}_n\underset{p}{\to}\boldsymbol{\beta}$，并且，

$$\sqrt{n}\,(\mathbf{x}_n-\boldsymbol{\beta})\underset{d}{\to}\mathbf{z}.$$

进一步，假设 $\mathbf{a}(\cdot):\mathbb{R}^K\to\mathbb{R}^r$ 一阶连续可微，其在 $\boldsymbol{\beta}$ 处的一阶导数为 $r\times K$ 维矩阵，记为 $\mathbf{A}(\boldsymbol{\beta})$，即，

$$\mathbf{A}(\boldsymbol{\beta})\equiv\frac{\partial\mathbf{a}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'},$$

则有：

$$\sqrt{n}\,[\mathbf{a}(\mathbf{x}_n)-\mathbf{a}(\boldsymbol{\beta})]\underset{d}{\to}\mathbf{A}(\boldsymbol{\beta})\mathbf{z}.$$

随机 O 和 o 符号

在渐近分析中，常引入一些特殊符号使推导过程得到简化，本小节介绍两种符号：O_p 和 o_p。

若随机向量序列 $\{\mathbf{x}_n,\,n\in\mathbb{N}\}$ 依概率收敛到 $\mathbf{0}$，则将其记为 $\mathbf{x}_n=o_p(1)$。在引入 O_p 之前，先定义何为依概率有界。具体地，令 $\{\mathbf{x}_n,\,n\in\mathbb{N}\}$ 为随机向量序列，若对任意的

$\varepsilon > 0$，存在常数 B 使得：

$$\sup_n \mathbb{P}(\|\mathbf{x}_n\| > B) < \varepsilon,$$

则称 $\{\mathbf{x}_n, n \in \mathbb{N}\}$ 依概率有界，记为 $\mathbf{x}_n = O_p(1)$。一般地，令 $\{\mathrm{x}_n\}$ 和 $\{\mathrm{y}_n\}$ 是两个随机变量的序列，符号 $\mathrm{y}_n = O_p(\mathrm{x}_n)$ 表示 $\{\mathrm{y}_n/\mathrm{x}_n\}$ 是 $O_p(1)$；符号 $\mathrm{y}_n = o_p(\mathrm{x}_n)$ 表示 $\mathrm{y}_n/\mathrm{x}_n \underset{p}{\to} 0$。

若随机向量序列 $\{\mathbf{y}_n\}$ 可以表示为 $\mathbf{y}_n = \mathbf{x}_n + o_p(1)$，则称 $\{\mathbf{y}_n\}$ 与 $\{\mathbf{x}_n\}$ 渐近等价，记为 $\mathbf{y}_n \underset{a}{\sim} \mathbf{x}_n$。

大数定律和中心极限定理

大数定律和中心极限定理分别涉及的是样本均值的概率极限和渐近分布，在样本独立同分布的情况下，常用的大数定律有：

- Khintchine 弱大数定律：令 $\{\mathrm{x}_i\}$ 满足独立同分布且具有有限期望 μ，则有：

$$\bar{\mathrm{x}}_n \equiv \frac{1}{n}\sum_{i=1}^{n} \mathrm{x}_i \underset{p}{\to} \mu.$$

- Kolmogorov 强大数定律：令 $\{\mathrm{x}_i\}$ 独立同分布，则有：

$$\bar{\mathrm{x}}_n \underset{a.s.}{\to} \mu \quad \text{当且仅当} \quad \mathbb{E}|\mathrm{x}_i| < \infty \text{ 且 } \mathbb{E}(\mathrm{x}_i) = \mu.$$

常用的中心极限定理有：

- Lindeberg-Levy 中心极限定理：令 $\{\mathbf{x}_i\}$ 满足独立同分布，$\mathbb{E}(\mathbf{x}_i) = \boldsymbol{\mu}$，$\mathrm{Var}(\mathbf{x}_i) = \boldsymbol{\Sigma}$，则有：

$$\sqrt{n}\,(\bar{\mathbf{x}}_n - \boldsymbol{\mu}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}) \underset{d}{\to} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

平滑函数模型

平滑函数模型是指 $\boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\mu})$，其中，$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{h}(\mathbf{x})]$，并且 $\boldsymbol{g}(\boldsymbol{\mu})$ 是平滑（即具有适当的连续性）函数。因为 $\boldsymbol{\theta}$ 不是总体矩，所以不能直接使用矩估计。估计 $\boldsymbol{g}(\boldsymbol{\mu})$ 的常用方法是先计算样本均值 $\hat{\boldsymbol{\mu}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{h}(\boldsymbol{x}_i)$，再将 $\hat{\boldsymbol{\mu}}$ 代入 $\boldsymbol{\theta}$ 的表达式得到 $\hat{\boldsymbol{\theta}} = \boldsymbol{g}(\hat{\boldsymbol{\mu}})$。许多估计量可以看作是平滑函数模型的特例，其中包括样本方差和最小二乘估计量。

下面的定理建立了平滑函数模型估计的一致性和渐近正态性：

定理 B.6. 平滑函数模型估计的渐近性质

令 $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ 是简单随机样本，$\boldsymbol{x}_i \in \mathbb{R}^m$，$\boldsymbol{h}(\cdot) : \mathbb{R}^m \to \mathbb{R}^k$，$\mathbb{E}\|\boldsymbol{h}(\mathbf{x})\| < \infty$，并且，$\boldsymbol{g}(\cdot) : \mathbb{R}^k \to \mathbb{R}^q$ 在 $\boldsymbol{\mu}$ 处连续，则有：

$$\hat{\boldsymbol{\theta}} \underset{p}{\to} \boldsymbol{\theta}, \quad n \to \infty.$$

若进一步，$\mathbb{E}\|\boldsymbol{h}(\mathbf{x})\|^2 < \infty$，$\mathbf{G}(\boldsymbol{u}) \equiv \dfrac{\partial}{\partial \boldsymbol{u}}\boldsymbol{g}(\boldsymbol{u})'$ 在 $\boldsymbol{\mu}$ 的邻域内连续，则有：

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \underset{d}{\to} \mathcal{N}(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\theta}}), \quad n \to \infty,$$

其中，$\boldsymbol{V}_{\boldsymbol{\theta}} = \mathbf{G}'\boldsymbol{V}\mathbf{G}$，$\boldsymbol{V} = \mathbb{E}[(\boldsymbol{h}(\mathbf{x}) - \boldsymbol{\mu})(\boldsymbol{h}(\mathbf{x}) - \boldsymbol{\mu})']$，$\mathbf{G} = \mathbf{G}(\boldsymbol{\mu})$。♡

时间序列的基本概念

为了定义时间序列，本小节先讨论随机过程。设 $(\boldsymbol{\Omega}, \mathscr{F}, \mathbb{P})$ 为一概率空间，T 为指标集，如果对于任意的 $t \in T$，均有定义在 $(\boldsymbol{\Omega}, \mathscr{F}, \mathbb{P})$ 上的随机变量 x_t，则集合 $\{\mathrm{x}_t,\ t \in T\}$ 称为随机过程。

根据随机过程的定义，对任意给定的 $t \in T$，x_t 是定义在 $\boldsymbol{\Omega}$ 上的函数；另一方面，若固定 $\omega \in \boldsymbol{\Omega}$，$\{\mathrm{x}_t(\omega),\ t \in T\}$ 是定义在 T 上的函数，此函数被称为随机过程 $\{\mathrm{x}_t,\ t \in T\}$ 的一次实现（realization），也称为 $\{\mathrm{x}_t,\ t \in T\}$ 的一条样本路径。在金融计量学中提及的时间序列数据是指某一随机过程的一次实现，例如，上证指数日收益率的历史数据实际上是日收益率[1]构成的随机过程的一次实现。虽然这里区分了随机过程和随机过程的一次实现，但是，通常在金融计量学中时间序列既可以指代随机过程，又可以指代随机过程的一次实现。

为了对时间序列数据进行统计分析，需要实现时间序列数据背后的随机过程满足类似于独立同分布的性质，为此，首先定义如下平稳性的概念，其在统计推断中的作用与 i.i.d. 中的同分布相似：

定义 B.2. 平稳性（Stationarity）

若随机过程 $\{\mathbf{x}_t\}$ 满足：对任意的 $r \in \mathbb{N}$ 和 $k \in \mathbb{Z}$，$(\mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_r})$ 和 $(\mathbf{x}_{t_1+k}, \ldots, \mathbf{x}_{t_r+k})$ 的联合分布相同，则称 $\{\mathbf{x}_t\}$ 为（严）平稳随机过程。

若 $\{\mathbf{x}_t\}$ 满足：

- $\mathbb{E}(\|\mathbf{x}_t\|^2) < \infty$；
- $\mathbb{E}(\mathbf{x}_t)$ 与 t 无关；
- $\mathrm{Cov}(\mathbf{x}_t, \mathbf{x}_{t-s})$ 存在且有限，其取值只与 s 有关，而与 t 无关，

则称 $\{\mathbf{x}_t\}$ 为弱平稳随机过程，也称为协方差平稳或二阶平稳随机过程。 ♣

从上述定义可以看出平稳性定义了随机过程在时间维度上的某种稳定性，严平稳是从联合分布的角度定义稳定性，而弱平稳是从一阶矩和二阶矩的角度定义稳定性。值得注意的是，严平稳并不一定蕴涵弱平稳，这是因为有些概率分布（如 Cauchy 分布）不存在有限的一阶矩和二阶矩。

时间序列分析中的一类重要弱平稳过程是如下定义的白噪声：

定义 B.3. 白噪声（White Noise, WN）

令 $\{\mathbf{x}_t\}$ 为弱平稳随机过程，如果：

$$\mathbb{E}(\mathbf{x}_t) = \mathbf{0}, \tag{B.9}$$

$$\mathrm{Cov}(\mathbf{x}_t, \mathbf{x}_{t-s}) = \mathbf{0}, \quad \forall\, s \neq 0, \tag{B.10}$$

则称 $\{\mathbf{x}_t\}$ 为白噪声，记为 $\mathbf{x}_t \sim \mathrm{WN}$。若将条件 (B.10) 替换为：对任意的 $s \neq 0$，$\mathbf{x}_t$ 和 $\mathbf{x}_{t-s}$ 相互独立，则称 $\{\mathbf{x}_t\}$ 为独立白噪声过程；若 $\{\mathbf{x}_t\}$ 满足条件 (B.10)，且对任意的 t，$\mathbf{x}_t$ 服从多元正态分布 $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$，则称 $\{\mathbf{x}_t\}$ 为高斯白噪声过程。 ♣

[1] 将日收益率理解为是随机变量。

除了时间维度上分布的稳定性，为了建立适用于时间序列的大数定律和中心极限定理还需要时间序列具有某种程度的独立性，为此引入如下遍历性的概念：

定义 B.4. 遍历性（Ergodicity）

令 $\{\mathrm{x}_t\}$ 为一平稳随机过程，如果对任意两个有界函数 $f:\mathbb{R}^{k+1}\to\mathbb{R}$ 和 $g:\mathbb{R}^{l+1}\to\mathbb{R}$，有：

$$\lim_{h\to\infty}\left|\mathbb{E}[f(\mathrm{x}_t,\ldots,\mathrm{x}_{t+k})g(\mathrm{x}_{t+h},\ldots,\mathrm{x}_{t+h+l})]\right| = \left|\mathbb{E}[f(\mathrm{x}_t,\ldots,\mathrm{x}_{t+k})]\right|\left|\mathbb{E}[g(\mathrm{x}_t,\ldots,\mathrm{x}_{t+l})]\right|,$$

则称 $\{\mathrm{x}_t\}$ 是遍历的，遍历的平稳随机过程称为遍历平稳过程。♣

从遍历性的定义可以看出遍历过程具有渐近的独立性。遍历平稳过程具有如下形式的大数定律：（证明见 Karlin et al. (1975) 的第九章定理 5.5）

定理 B.7. 遍历定理（Ergodic Theorem）

令 $\{\mathbf{x}_t\}$ 是遍历平稳过程，$\mathbb{E}(\mathbf{x}_t)=\boldsymbol{\mu}$，则有：

$$\bar{\mathbf{x}}_T \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \underset{a.s.}{\to} \boldsymbol{\mu}.$$

♡

为了叙述本书使用的遍历平稳时间序列的中心极限定理，需要引入鞅和鞅差序列的概念：令 $\{\mathbf{y}_t,\ t\in T\}$ 和 $\{\mathbf{x}_t,\ t\in T\}$ 为两个随机过程，如果：

- $\mathbf{y}_t$ 是 $\mathbf{x}_1,\ldots,\mathbf{x}_t$ 的（可测）函数；
- $\mathbb{E}[\|\mathbf{y}_t\|]<\infty$，$\forall\, t$；
- $\mathbb{E}[\mathbf{y}_{t+1}\mid \mathbf{x}_1,\ldots,\mathbf{x}_t]=\mathbf{y}_t$，$\forall\, t$，

则称 $\{\mathbf{y}_t,\ t\in T\}$ 是关于 $\{\mathbf{x}_t,\ t\in T\}$ 的鞅（martingale）序列。若 $\{\mathbf{g}_t,\ t\in T\}$ 满足：

- $\mathbf{g}_t$ 是 $\mathbf{x}_1,\ldots,\mathbf{x}_t$ 的（可测）函数；
- $\mathbb{E}[\|\mathbf{g}_t\|]<\infty$，$\forall\, t$；
- $\mathbb{E}[\mathbf{g}_{t+1}\mid \mathbf{x}_1,\ldots,\mathbf{x}_t]=\mathbf{0}$，$\forall\, t$，

则称 $\{\mathbf{g}_t,\ t\in T\}$ 是关于 $\{\mathbf{x}_t,\ t\in T\}$ 的鞅差序列（martingale difference sequence, MDS）。

若 $\{\mathbf{y}_t,\ t\in T\}$ 是关于 $\{\mathbf{x}_t,\ t\in T\}$ 的鞅序列，定义：

$$\mathbf{g}_t=\mathbf{y}_t-\mathbf{y}_{t-1},$$

则有：

$$\mathbb{E}[\mathbf{g}_{t+1}\mid \mathbf{x}_1,\ldots,\mathbf{x}_t]=\mathbb{E}[\mathbf{y}_{t+1}-\mathbf{y}_t\mid \mathbf{x}_1,\ldots,\mathbf{x}_t]=\mathbf{0}.$$

因此，鞅序列的一阶差分序列即为鞅差序列。若 $\{\mathbf{y}_t,\ t\in T\}$ 是关于其自身的鞅序列，则简称 $\{\mathbf{y}_t,\ t\in T\}$ 是鞅序列；类似地，若 $\{\mathbf{g}_t,\ t\in T\}$ 是关于其自身的鞅差序列，则简称 $\{\mathbf{g}_t,\ t\in T\}$ 是鞅差序列。

鞅差序列的一个重要性质是它不存在序列相关。构建金融时间序列模型通常需要不存在序列相关的零均值弱平稳过程作为基石。到目前为止，按照由强到弱的顺序，此类过程可以设置为：独立白噪声过程 $\Rightarrow$ 具有有限方差的平稳鞅差序列 $\Rightarrow$ 白噪声。

下面的中心极限定理将 Lindeberg-Levy 中心极限定理推广到遍历平稳的 MDS：（证明见 Billingsley (1961)）

定理 B.8. 遍历平稳鞅差序列的中心极限定理

令 $\{\mathbf{g}_t\}$ 是遍历平稳的向量鞅差序列，满足 $\mathbb{E}(\mathbf{g}_t\mathbf{g}_t') = \mathbf{\Sigma}$，定义：

$$\bar{\mathbf{g}} \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t,$$

则有：

$$\sqrt{T}\,\bar{\mathbf{g}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_t \underset{d}{\to} \mathcal{N}(\mathbf{0},\mathbf{\Sigma}).$$

B.5 Bootstrap 的渐近理论

令 $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ 是满足独立同分布的 n 个随机向量，$\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}$ 是这 n 个随机向量的一组样本观测。对 $\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}$ 进行 n 次有放回的随机抽样得到 Bootstrap 样本 $\{\boldsymbol{x}_1^*,\ldots,\boldsymbol{x}_n^*\}$。注意到 Bootstrap 样本的样本均值 $\bar{\boldsymbol{x}}_n^* \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^*$ 也是 n 个独立同分布的随机变量的样本观测值的平均，直观上，它应该依概率收敛到 $\mathbb{E}(\mathbf{x}_i)$。但是，Bootstrap 样本均值的渐近性质涉及到条件分布，这是因为 Bootstrap 是以给定的样本 $\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}$ 为条件。为了区别于前文定义的依概率收敛，下面定义条件分布的依概率收敛：令 $\{\mathbf{z}_n^*,\, n\in\mathbb{N}\}$ 是随机向量序列，$\mathbf{z}$ 为一随机向量，如果，对任意的 $\varepsilon>0$，有：

$$\mathbb{P}^*\left[\|\mathbf{z}_n^*-\mathbf{z}\|>\varepsilon\right]\underset{p}{\to}0,\quad n\to\infty$$

其中，$\mathbb{P}^*[\cdot]$ 表示条件概率，则称 $\{\mathbf{z}_n^*\}$ 依 Bootstrap 概率收敛到 $\mathbf{z}$，记为 $\mathbf{z}_n^*\underset{p^*}{\to}\mathbf{z}$。容易验证：如果 $\mathbf{z}_n\underset{p}{\to}\mathbf{z}$，则 $\mathbf{z}_n\underset{p^*}{\to}\mathbf{z}$。

除了依 Bootstrap 概率收敛，还可以定义依 Bootstrap 分布收敛：令 $\{\mathbf{z}_n^*,\, n\in\mathbb{N}\}$ 是随机向量序列，其条件分布为 $G_n^*(\boldsymbol{x})=\mathbb{P}^*[\mathbf{z}_n^*\leqslant\boldsymbol{x}]$。如果，对所有的 $\boldsymbol{x}$ 使得 $G(\boldsymbol{x})=\mathbb{P}[\mathbf{z}\leqslant\boldsymbol{x}]$ 连续，有 $G_n^*(\boldsymbol{x})\underset{p}{\to}G(\boldsymbol{x})$，则称 $\{\mathbf{z}_n^*\}$ 依 Bootstrap 分布收敛到 $\mathbf{z}$，记为 $\mathbf{z}_n^*\underset{d^*}{\to}\mathbf{z}$。

对于 Bootstrap 样本均值，有如下的 Bootstrap 弱大数定律和中心极限定理：

- Bootstrap 弱大数定律：如果 $\mathbf{x}_i, i=1,\ldots,n$，相互独立且一致可积[2]，则有：$\bar{\boldsymbol{x}}_n^*-\bar{\boldsymbol{x}}_n\underset{p^*}{\to}\mathbf{0}$，$\bar{\boldsymbol{x}}_n^*\underset{p^*}{\to}\boldsymbol{\mu}=\mathbb{E}[\mathbf{x}_i]$，其中，$\bar{\boldsymbol{x}}_n$ 为原始样本的样本均值；
- Bootstrap 中心极限定理：如果 $\mathbf{x}_i$，$i=1,\ldots,n$，是 i.i.d. 的，$\mathbb{E}\|\mathbf{x}_i\|^2<\infty$，且 $\mathbf{\Sigma}=\mathrm{Var}[\mathbf{x}_i]$ 正定，则有：$\sqrt{n}\,(\bar{\boldsymbol{x}}_n^*-\bar{\boldsymbol{x}}_n)\underset{d^*}{\to}\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$。

连续映射定理和 Delta 方法也有相应的 Bootstrap 版本：

定理 B.9. Bootstrap 连续映射定理

- 如果 $\mathbf{z}_n^*\underset{p^*}{\to}\boldsymbol{c}$ 且 $\boldsymbol{g}(\cdot)$ 在 $\boldsymbol{c}$ 点连续，则有 $\boldsymbol{g}(\mathbf{z}_n^*)\underset{p^*}{\to}\boldsymbol{g}(\boldsymbol{c})$；
- 如果 $\mathbf{z}_n^*\underset{d^*}{\to}\mathbf{z}$，并且 $\boldsymbol{g}:\mathbb{R}^m\to\mathbb{R}^k$ 的不连续点构成的集合 $D_{\boldsymbol{g}}$ 满足 $\mathbb{P}^*[\mathbf{z}_n^*\in$

[2] 令 $\mathbf{z}_n$ 是随机向量，如果

$$\lim_{M\to\infty}\limsup_{n\to\infty}\mathbb{E}\left[\|\mathbf{z}_n\|\,\mathbf{1}\{\|\mathbf{z}_n\|>M\}\right]=0,\quad n\to\infty,$$

则称 $\mathbf{z}_n$ 一致可积。

$D_{\boldsymbol{g}}] = 0$，则有 $\boldsymbol{g}(\mathbf{z}_n^*) \underset{d^*}{\rightarrow} \boldsymbol{g}(\mathbf{z})$。

定理 B.10. Bootstrap Delta 方法

如果 $\hat{\boldsymbol{\mu}} \underset{p}{\rightarrow} \boldsymbol{\mu}$，$\sqrt{n}\left(\hat{\boldsymbol{\mu}}^* - \hat{\boldsymbol{\mu}}\right) \underset{d^*}{\rightarrow} \xi$，并且 $\boldsymbol{g}(\cdot)$ 在 $\boldsymbol{\mu}$ 的邻域内连续可微，则有：

$$\sqrt{n}\left(\boldsymbol{g}\left(\hat{\boldsymbol{\mu}}^*\right) - \boldsymbol{g}(\hat{\boldsymbol{\mu}})\right) \underset{d^*}{\rightarrow} \mathbf{G}'\xi,$$

其中，$\mathbf{G}(\boldsymbol{x}) = \dfrac{\partial}{\partial \boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x})'$，$\mathbf{G} = \mathbf{G}(\boldsymbol{\mu})$。

前文提到金融计量学中的许多估计量都是平滑函数模型的特例，可以证明平滑函数模型具有如下性质：

定理 B.11. 平滑函数模型估计的 Bootstrap 渐近性质

设定理 B.6 的条件成立，并且，$\hat{\boldsymbol{\theta}}^* = \boldsymbol{g}(\hat{\boldsymbol{\mu}}^*)$，其中，$\hat{\boldsymbol{\mu}}^* = n^{-1}\sum_{i=1}^n \boldsymbol{h}(\boldsymbol{x}_i^*)$，则有；

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}\right) \underset{d^*}{\rightarrow} \mathcal{N}(\mathbf{0},\, \boldsymbol{V}_{\boldsymbol{\theta}}), \quad n \rightarrow \infty,$$

其中，$\boldsymbol{V}_{\boldsymbol{\theta}} = \mathbf{G}'\boldsymbol{V}\mathbf{G}$，$\boldsymbol{V} = \mathbb{E}[(\boldsymbol{h}(\mathbf{x}) - \boldsymbol{\mu})(\boldsymbol{h}(\mathbf{x}) - \boldsymbol{\mu})']$，$\mathbf{G} = \mathbf{G}(\boldsymbol{\mu})$。

此结论表明 Bootstrap 估计量 $\hat{\boldsymbol{\theta}}^*$ 的渐近分布与 $\hat{\boldsymbol{\theta}}$ 的渐近分布完全相同，可以通过 Bootstrap 分布获取与 $\hat{\boldsymbol{\theta}}$ 的分布相关的信息。因此，利用 Bootstrap 分布可以进行渐近正确（correct）的统计推断。

本节列出的大部分结论的证明都可以在 Hansen (2022a) 的第 10.31 节中找到。

B.6 多元正态分布及其性质

令 $\mathbf{x}$ 是 m 维随机向量，具有联合概率密度函数：

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{m/2}(\det \boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}, \tag{B.11}$$

其中，$\boldsymbol{\mu}$ 是 m 维常数向量，$\boldsymbol{\Sigma}$ 是 m 阶正定矩阵，则称 $\mathbf{x}$ 服从均值为 $\boldsymbol{\mu}$，协方差矩阵为 $\boldsymbol{\Sigma}$ 的 m 元正态分布，记为 $\mathbf{x} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$。

多元正态分布具有如下常用的性质：（证明请参考 Johnson et al. (2007) 的第四章或李高荣，吴密霞 (2021) 的第三章）

1. 如果 $\mathbf{x} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则其任意线性组合 $\boldsymbol{a}'\mathbf{x}$ 服从一元正态分布 $\mathcal{N}(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$；反之，若对任意的 $\boldsymbol{a}$，$\boldsymbol{a}'\mathbf{x} \sim \mathcal{N}(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$，则 $\mathbf{x}$ 一定服从多元正态分布 $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$；
2. 如果 $\mathbf{x} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，$\boldsymbol{A}$ 是 $n \times m$ 维常数矩阵，则 $\boldsymbol{A}\mathbf{x} \sim \mathcal{N}_n(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$；
3. 如果 $\mathbf{x} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，$\boldsymbol{d}$ 是 m 维常数向量，则 $\mathbf{x} + \boldsymbol{d} \sim \mathcal{N}_m(\boldsymbol{\mu} + \boldsymbol{d}, \boldsymbol{\Sigma})$；
4. 如果 $\mathbf{x}$ 服从多元正态分布，则 $\mathbf{x}$ 的子集也服从多元正态分布。具体地，将 $\mathbf{x}$、$\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 按照如下方式进行分割：

$$\underset{(m\times 1)}{\mathbf{x}} = \begin{bmatrix} \underset{(q_1\times 1)}{\mathbf{x}_1} \\ \underset{(q_2\times 1)}{\mathbf{x}_2} \end{bmatrix}, \quad \underset{(m\times 1)}{\boldsymbol{\mu}} = \begin{bmatrix} \underset{(q_1\times 1)}{\boldsymbol{\mu}_1} \\ \underset{(q_2\times 1)}{\boldsymbol{\mu}_2} \end{bmatrix} \tag{B.12}$$

和

$$\underset{(m\times m)}{\boldsymbol{\Sigma}} = \begin{bmatrix} \underset{(q_1\times q_1)}{\boldsymbol{\Sigma}_{11}} & \underset{(q_1\times q_2)}{\boldsymbol{\Sigma}_{12}} \\ \underset{(q_2\times q_1)}{\boldsymbol{\Sigma}_{21}} & \underset{(q_2\times q_2)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}, \tag{B.13}$$

则有：$\mathbf{x}_1 \sim \mathcal{N}_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$，$\mathbf{x}_2 \sim \mathcal{N}_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$；

5. 如果 $\mathbf{x} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，式 (B.12) 中的 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 相互独立当且仅当式 (B.13) 中的 $\boldsymbol{\Sigma}_{12} = \mathbf{0}$；
6. 如果 $\mathbf{x}_1 \sim \mathcal{N}_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$，$\mathbf{x}_2 \sim \mathcal{N}_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$，并且，$\mathbf{x}_1$ 和 $\mathbf{x}_2$ 相互独立，则 $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2')'$ 服从多元正态分布，其期望为 $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')'$，协方差矩阵为由 $\boldsymbol{\Sigma}_{11}$ 和 $\boldsymbol{\Sigma}_{22}$ 构成的分块对角阵：

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{22} \end{bmatrix};$$

7. 设 $\mathbf{x}_1, \ldots, \mathbf{x}_k$ 是相互独立的 k 个 m 维随机向量，如果对任意的 $1 \leqslant i \leqslant k$，$\mathbf{x}_i \sim \mathcal{N}_m(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$，则有：

$$\sum_{i=1}^{k} \mathbf{x}_i \sim \mathcal{N}_m\left(\sum_{i=1}^{k} \boldsymbol{\mu}_i, \sum_{i=1}^{k} \boldsymbol{\Sigma}_i\right).$$

在金融计量学中，许多检验统计量（渐近）分布的推导需要用到如下关于二次型的概率分布的结论：

定理 B.12. 二次型的概率分布

令 $\mathbf{x}$ 是 m 维随机向量，如果 $\mathbf{x} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I}_m)$，并且，$\mathbf{A}$ 是对称幂等矩阵，则有 $\mathbf{x}'\boldsymbol{A}\mathbf{x}$ 服从卡方分布，其自由度为 $\mathrm{rank}(\boldsymbol{A})$。♡

定理 B.13. 平方马氏距离的概率分布

令 $\mathbf{x}$ 是 m 维随机向量，如果 $\mathbf{x} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，并且，$\boldsymbol{\Sigma}$ 非奇异，则有 $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \sim \chi^2(m)$. ♡

本节的最后给出多元正态分布的条件分布性质，它在卡尔曼滤波等金融计量方法中起到关键的作用：

定理 B.14. 多元正态分布的条件分布

令 $\mathbf{x} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，将 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 按照 (B.12) 和 (B.13) 的方式分割，

- 若 $\boldsymbol{\Sigma}_{22}$ 非奇异，则以 $\mathbf{x}_2 = \boldsymbol{x}_2$ 为条件，$\mathbf{x}_1$ 的条件分布为：

$$\mathbf{x}_1 \mid \mathbf{x}_2 = \boldsymbol{x}_2 \sim \mathcal{N}_{q_1}(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}), \tag{B.14}$$

  其中，

$$\begin{aligned} \boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \end{aligned}$$

- 若 $\mathbf{\Sigma}_{11}$ 非奇异，则以 $\mathbf{x}_1 = \boldsymbol{x}_1$ 为条件，$\mathbf{x}_2$ 的条件分布为：

$$\mathbf{x}_2 \mid \mathbf{x}_1 = \boldsymbol{x}_1 \sim \mathcal{N}_{q_2}(\boldsymbol{\mu}_{2|1}, \mathbf{\Sigma}_{2|1}), \tag{B.15}$$

其中，

$$\begin{aligned} \boldsymbol{\mu}_{2|1} &= \boldsymbol{\mu}_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \\ \mathbf{\Sigma}_{2|1} &= \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}. \end{aligned}$$

此定理的证明请参考 Johnson et al. (2007) 的 Result 4.6。

B.7 常用不等式

- Jensen 不等式：对任意的随机变量 $\mathbf{x}$，如果 $g(x)$ 是凸函数，则：

$$g(\mathbb{E}[\mathbf{x}]) \leqslant \mathbb{E}[g(\mathbf{x})];$$

  若 $g(x)$ 是凹函数，则：

$$\mathbb{E}[g(\mathbf{x})] \leqslant g(\mathbb{E}[\mathbf{x}]).$$

- 期望不等式：对任意的随机向量 $\mathbf{y} \in \mathbb{R}^n$，$\mathbb{E}\|\mathbf{y}\| < \infty$，有：

$$\|\mathbb{E}[\mathbf{y}]\| \leqslant \mathbb{E}\|\mathbf{y}\|. \tag{B.16}$$

  进一步，对任意的随机矩阵 $\mathbf{Y} \in \mathbb{R}^{m\times n}$，$\mathbb{E}\|\mathbf{Y}\| < \infty$，有：

$$\|\mathbb{E}[\mathbf{Y}]\| \leqslant \mathbb{E}\|\mathbf{Y}\|. \tag{B.17}$$

- Chebyshev 不等式：对任意的随机变量 $\mathrm{x} \in \mathbb{R}$ 以及任意的 $\varepsilon > 0$，

$$\mathbb{P}\left[|\mathrm{x} - \mathbb{E}(\mathrm{x})| > \varepsilon\right] \leqslant \frac{\mathrm{Var}[\mathrm{x}]}{\varepsilon^{-2}}.$$

- 一般形式的 Chebyshev 不等式：令 x 为随机变量，$g(x) > 0$ 是定义在 $\mathbb{R}$ 上的非递减函数，则有，对任意的 ε，

$$\mathbb{P}(\mathrm{x} \geqslant \varepsilon) \leqslant \frac{\mathbb{E}[g(\mathrm{x})]}{g(\varepsilon)}.$$

- Hölder 不等式：对任意的 $p > 1$ 和 $q > 1$ 使得 $\frac{1}{p} + \frac{1}{q} = 1$，以及任意的 $m \times n$ 维随机矩阵 $\mathbf{X}$ 和 $\mathbf{Y}$，

$$\mathbb{E}\left\|\mathbf{X}'\mathbf{Y}\right\| \leqslant \left(\mathbb{E}\|\mathbf{X}\|^p\right)^{1/p}\left(\mathbb{E}\|\mathbf{Y}\|^q\right)^{1/q}.$$